

## Evaluating the Performance of Autoregressive Model for Solar Radiation Forecasting

Ocholi M.<sup>1,\*</sup>, Adeyemi, B.<sup>2</sup>, Omojola O. O.<sup>3</sup> and Samuel C. S.<sup>4</sup>

<sup>1,3,4</sup>Science Laboratory and Technology Research Department, Nigerian Building and Road Research Institute, P.M.B 1055, Ota, Nigeria

<sup>2</sup>Department of Physics, Federal University of Technology Akure, Nigeria

Corresponding Author: \*mamuduocholi@nbrri.gov.ng

<https://doi.org/10.36263/nijest.2021.02.0289>

### ABSTRACT

The solar radiation data taken from 14 meteorological stations in Nigeria has been analyzed. The periodic component of the data which covered a period of 13 (mostly 1977-1989) years was removed via Fourier analysis while the residual series was subjected to autoregressive analysis. It was evident from the t-test and autocorrelation plots of the modified (i.e. without the periodic component) series that there exist significant persistence at nine stations including Sokoto, Nguru, Kano, Maiduguri, Bauchi, Yola, Minna, Ibadan, and Benin. The autocorrelation at Jos, Bida, Ikeja, Enugu and Port Harcourt were however found to be insignificant. As the sample partial autocorrelation function cuts off after lag 1, a non-seasonal autoregressive model of order 1, AR (1), was identified for stations with autocorrelation. The Q-statistic of error series suggested that the models were adequate as identified. Moreover, the exploratory plots of the model residual series showed agreement with the quantitative statistics and thus enforces the inference that the models were adequate for monthly mean daily global solar radiation forecasts at some of the study stations. It is interesting to note that all the stations within the sub-sahelian region showed significant persistence whereas all the stations in the coastal region except Benin were found with insignificant autocorrelation. Expectedly, the performance evaluation of the model gave impressive result for the stations within the sub-sahelian region but a relatively weak result for the coastal region. The result for the midland region was mixed whereas it was difficult to conclude on the Guinea savannah region with result from only one station.

**Keywords:** Modelling, Forecasting, Solar radiation, Autocorrelation, Autoregressive model

### 1.0. Introduction

Solar radiation is the radiant energy flux emanating from the sun. As it travels through space to the top of the Earth's atmosphere, it is relatively unaffected in spectral distribution and intensity. However, on its way to the Earth's surface, it is subjected to scattering and absorption by atmospheric constituents. Studies have shown that some 25-70% of solar radiation reaching the top of the atmosphere is received at the Earth's surface. Thus, seasonal variation in the composition and concentration of atmospheric constituents determines to a great extent the spatial and temporal variability of solar radiation at the Earth's surface (Brooks, 2006; Lindsey, 2009).

In research and application fields such as architecture, energy, agriculture, meteorology to mention a few, it is usually required that the variable be measured continuously and accurately over a long term. However, there is a huge cost implication to continuously measure solar radiation at all locations. Consequently, many studies have been carried out to develop methods to estimate climatic variables such as solar radiation at places where measurements are not available but share similar climatic conditions with the measured locations. Some researchers capitalize on the ability of individual variable to predict self (persistence), others rely on other measured parameters known as explanatory variables in establishing empirical regression models.

Numerous empirical studies have been carried out to investigate the predictive skill of persistence in meteorological and geophysical time series. For instance, in the study of Sulaiman *et al.* (1997), the daily solar radiation data from four different locations in Malaysia was analyzed using the Box-Jenkins approach. The deterministic component is removed using Fourier analysis and the stochastic component was subjected to Autoregressive Moving Average (ARMA) analysis. Similarly, Zakaria (2011) developed a model for the periodic and stochastic components of monthly rainfall data from Purajaya station, Indonesia and was able to adequately synthesize monthly rainfall data for the Purajaya area. Zaharim *et al.* (2009) in their work applied the Box-Jenkins method to solar radiation data from Bangi, Malaysia. The non-seasonal autoregressive model of order 1, ARMA (1, 0), was found to model the data adequately. In an insightful work of Barnett *et al.* (1984) where they investigated the role of persistence in the predictability of short-term climate variation over Eurasian continent, it was found that persistence in air temperature anomalies accounted for most of the predictability ahead of sea surface temperature and sea level pressure anomalies as other predictors. Adeyemi and Fasoranbaku (2008) sought to predict daily mean air temperature over Nigeria and developed non-seasonal autoregressive models of varying order for eight meteorological sites. Results obtained were found to compare well with existing models.

Some other authors have fit seasonal ARMA (SARMA) model and periodic autoregressive (PAR) model to meteorological time series in order to explicitly account for seasonal variation. For instance, Tesfaye *et al.* (2006) developed model identification and simulation techniques based on the periodic autoregressive moving average (PARMA) model that could capture the seasonal variations in river flow statistics. The technique was applied to monthly flow data for the Fraser River in British Columbia and it was found that the statistical analysis of the PARMA model residuals, including a truncated Pareto model for the extreme tails produced a realistic simulation of these river flows. Smadi (2009) investigated the best of the traditional SARMA model and the relatively new PAR model for modeling the average monthly maximum temperature data in Jordan. He found that the PAR model does a better job but the fact that large number of parameters is involved, especially for monthly data, remained its main drawback.

Besides, meteorological time series can also be estimated based on other measured variables. For instance solar radiation has been estimated based on variables such as sunshine hours (Page, 1964; Akpabio *et al.*, 2004; Falayi and Rabiu, 2005; Safari and Gasore, 2009), air temperature (Okundamiya and Nzeako, 2011), precipitation and relative humidity (Trabea and Shaltout, 2000). The most widely used model in those studies is the Angstrom-Prescott model.

In this paper, we sought to develop persistence models for monthly mean daily global solar radiation forecast over Nigeria. This specific effort looked fruitful since prior studies have suggested the existence of strong persistence in global solar radiation elsewhere within the tropical belt (Zaharim *et al.*, 2009; Raji *et al.*, 2012).

## 2.0. Methodology

### 2.1. Data sourcing and cleaning

The data for this study were obtained from the Nigerian Meteorological Agency (NIMET), Federal Ministry of Aviation, Oshodi, Lagos. A 13-year monthly mean daily global solar radiation data from fourteen meteorological stations in Nigeria was acquired. The first 8-year data was used to calibrate the data and the remaining 5-year data was used to validate the model.

In most of the stations it was found that the data is filled with missing measurements. The long term mean of the month to which the missing measurement belongs was used to replace the missing values in the data set. The data was also inspected for the presence of outliers via the notched box plots. The values lying beyond the whiskers are marked as outliers in comparison to what is expected from a normal distribution with the same mean and variance as the sample data. Some of the stations including Sokoto, Nguru, Maiduguri, Bauchi, Jos, Bida, Ibadan, Ikeja, Enugu and Benin were found with a number of outliers. This is especially important since the least squares method of model

parameter estimation employed in this study is very sensitive to the presence of outliers and could lead to spurious results if not checked. The spotted outliers were removed and replaced by the long term mean of the month to which they belong in the data set.

## 2.2. Deterministic and stochastic components

Solar radiation, like most meteorological variables consists of periodic component (deterministic) and random component (non-deterministic). The periodic component is predictable with great precision even far into the future but it is not so with the random component (Jones, 1964). It suffices then that when doing a statistical analysis of a meteorological time series; it should be the aim to determine the predictability of this nondeterministic component. Thus, suppose  $x_0, x_1, \dots, x_{n-1}$  is an arbitrary time series of length  $n$ , the time series can be expressed as (Sharma, 1989):

$$x_t = P_t + r_t \quad t = 1, 2, \dots, n-1 \quad (1)$$

where,  $P_t$  is the periodic component and  $r_t$  is the stochastic or random component of the series.

The periodic component can be expressed as the sum of sinusoids at the dominant Fourier frequencies of the series as (Sharma, 1989; Zakaria, 2011):

$$P_t = \sum_{j=0}^k [a_j \cos(2\pi f_j t) + b_j \sin(2\pi f_j t)], \quad (2)$$

The summation in (2) is over the Fourier frequencies given by:

$$f_j = \frac{j}{n}, \quad j = 0, 1, 2, k. \quad (3)$$

$f_j$  are the harmonics,  $a_j$  and  $b_j$  are the Fourier coefficients. The dominant harmonics are obtained from the spectral estimates of the time series (Sharma, 1989). In particular for the study stations, it was found that the first two harmonics are significant, implying that  $k=2$  for our time series. Thus the time distribution of the series may be written as:

$$x_t = \sum_{j=0}^2 [a_j \cos(2\pi f_j t) + b_j \sin(2\pi f_j t)] + r_t \quad (4)$$

where the term  $\sum_{j=0}^2 [a_j \cos(2\pi f_j t) + b_j \sin(2\pi f_j t)]$  is the estimated value of the periodic component and is denoted by  $\hat{P}_t$ .

The deterministic and stochastic components can be separated by subtracting the estimated periodic component  $\hat{P}_t$  month by month and dividing it by the sample standard deviation,  $s_t$  for same month (Sharma, 1989; Adamowski and Smith, 1972). Thus:

$$r_t = \frac{x_t - \hat{P}_t}{s_t} \quad (5)$$

where  $r_t$  is the stochastic component with zero mean and unit variance. The stochastic component may in turn be further decomposed to:

$$r_t = d_t + e_t, \quad (6)$$

where

$$d_t = \sum_{i=1}^p \phi_i r_{t-i}, \quad t = 1, 2, \dots, n \quad (7)$$

is the deterministic component representing persistence or autocorrelation in the residual series,  $\phi_i$  is the autocorrelation coefficients and  $e_t$  is white noise which is independently distributed with zero mean and constant variance and assumed to approximate a normal probability distribution (Sharma, 1989 ; Zakaria, 2011).

## 2.3. Stochastic modelling

The autoregressive model as in all other persistence models describes a stationary process. Consequently, a preliminary step in any time series modeling is to investigate if the underlying process generating the data is stationary. If the series is non-stationary, it must be transformed to

stationary series. Differencing is commonly used to transform a non-stationary time series into a stationary series. Stationary time series can be identified by conducting the unit root test on the series. The common test statistics for unit root are the Phillips-Perron test (PP test) and the Augmented Dickey-Fuller test (ADF test) which is basically an extension of the Dickey-Fuller test (DF test). These statistics evaluates the null hypothesis of the presence of a unit root for the model underlying a series against the alternative hypothesis of a root smaller than one. Suppose that  $r_t$  is a first order autoregressive process, the objective of the tests is to evaluate the hypotheses:

$$\left. \begin{array}{l} H_0: \phi_1 = 1 \text{ (i.e. series contains a unit root)} \\ H_1: \phi_1 < 1 \text{ (i.e. series is stationary)} \end{array} \right\} \quad (8)$$

For the underlying model given by:

$$r_t = \phi_1 r_{t-1} + e_t \quad (9)$$

The decision rule is that:

$$\left. \begin{array}{ll} \text{If } pp \text{ value} < C \text{ value}_{\alpha=0.05} & \text{Reject } H_0 \\ \text{If } pp \text{ value} > C \text{ value}_{\alpha=0.05} & \text{Not Reject } H_0 \end{array} \right\} \quad (10)$$

where *pp value* is the computed statistic, say the Phillips-Perron statistic at lag  $j$  and *C value* is the critical value which does not follow the usual t-distribution but follows a non-standard distribution and derived from Monte Carlo experiments (Fuller, 1976).

Most standard statistical software including MATLAB which is used in this study presents a table of critical values given any level of significance. Where the computed test statistic is more negative than the critical value, the null hypothesis of unit root is rejected implying a stationary series.

A complementary approach is to examine the behavior of the correlograms for the sample autocorrelation function (SACF) and the partial autocorrelation function (PACF) of the time series for several lags (say 20 lags). The sample autocorrelation function  $\rho_k$  at lag  $k$  is defined as:

$$\rho_k = \frac{\sum_{t=1}^{n-k} (r_t - \bar{r})(r_{t+k} - \bar{r})}{\sum_{t=1}^n (r_t - \bar{r})^2} \quad (11)$$

The plot of the sample autocorrelation function (SACF) as a function of lag is called the correlogram.  $\bar{r}$  is defined as the mean of the series,  $r_t$ .

The sample partial autocorrelation function (SPACF),  $\rho_k^p$  at lag  $k$  is defined as (Durbin, 1960):

$$\left. \begin{array}{l} \rho_k^p = \frac{\rho_k - \sum_{j=1}^{k-1} \rho_{j,k-1}^p * \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \rho_{j,k-1}^p * \rho_j} \\ \text{and} \\ \rho_{j,k}^p = \rho_{j,k-1}^p - \rho_{k,k}^p * \rho_{k-j,k-1}^p \end{array} \right\} \quad (12)$$

In order to establish the significance of the ACF values at various lags and for each station, 95% confidence bounds may be computed and superimposed on the correlograms for easy identification. In general, if the SACF of the time series either cuts off or dies down quickly, the time series should be considered stationary and conversely if the SACF values dies down extremely slowly, the series is considered non-stationary (Bowerman *et al.*, 2005). The presence of a trend and/or a periodic signal could cause a violation of the condition of stationarity. A common approach to removing trend and periodic signal in the time series is by de-trending and harmonic analysis respectively (Sharma, 1989; Zakaria, 2011).

The structure and order of the model is determined by examining the behavior of the SACF and the sample partial autocorrelation function (SPACF). If the SACF dies down and the SPACF cuts off after lag  $p$ , then an autoregressive process of order  $p$  (Box and Jenkins, 1976; Chatfield, 2004) is indicated and the model, in general is given as:

$$r_t = \emptyset_1 r_{t-1} + \emptyset_2 r_{t-2} + \dots + \emptyset_p r_{t-p} + e_t \quad (13)$$

The historic data is then fitted to the identified model by least squares method. We can judge the importance of the autocorrelation parameters using t-test. In doing this we can test the null hypothesis  $H_0$  against the alternative hypothesis  $H_1$  as follows:

$$\left. \begin{array}{l} H_0: \hat{\emptyset}_j = 0 \\ H_1: \hat{\emptyset}_j \neq 0 \end{array} \right\} \quad (14)$$

where,  $\hat{\emptyset}_j$  are the autoregressive parameter estimates from lag 1 to  $j$ . Thus we establish the decision rule as:

$$\left. \begin{array}{l} \text{if } |t_{\hat{\emptyset}_j}| \geq t_{n-k-1}^{\alpha/2} \text{ Reject } H_0 \\ \text{if } |t_{\hat{\emptyset}_j}| < t_{n-k-1}^{\alpha/2} \text{ Not Reject } H_0 \end{array} \right\} \quad (15)$$

where,  $t_{n-k-1}^{\alpha/2}$  is the value from the student's  $t$  distribution table at half the preset significance level  $\alpha$  (since it is a two-tail test) and at  $n - k - 1$  degrees of freedom.

Possible values of  $\alpha$  include 0.10, 0.05 and 0.01,  $n$  is the sample size and  $k$  is the number of parameter estimators. Thus, a rejection of the null hypothesis would suggest that the parameter estimate in question is statistically significant.

After establishing the model, we then check the adequacy of the model by examining the ‘whiteness’ of model residuals. This is of key importance in showing the effectiveness of the model in monitoring the underlying autoregressive process. The primary tool for model diagnosis is residual analysis. If the model adequately describes the time series, then the residual chronologies should be independent of one another as effective autocorrelation model is expected to remove autocorrelation in the time series leaving a random series. In testing randomness of the residual, the portmanteau test such as the one proposed by Ljung and Box (1978) is carried out on the residuals. The Ljung-Box statistic is also known as the Q-statistic and is given by:

$$Q = n'(n' + 2) \sum_{l=1}^k \frac{r_l^2(\hat{e})}{(n'-l)^{-1}} \quad (16)$$

with  $n' = n - d$ , where  $n$  is the number of observations in the series,  $d$  is the degree of differencing and  $r_l^2(\hat{e})$  is the sample autocorrelation of the residual at lag  $l$ .

The Q-statistic, computed from the lowest  $k$  autocorrelations, say at 1,2,3, . . . ,20 follows a  $\chi^2$  distribution with  $(k-p-q)$  degrees of freedom, where  $p$  and  $q$  are the AR and MA orders of the model. The Q-statistic is used to evaluate the following hypothesis:

$$\left. \begin{array}{l} H_0: \hat{\emptyset}_k(\hat{e}) = 0 \\ H_1: \hat{\emptyset}_k(\hat{e}) \neq 0 \end{array} \right\} \quad (17)$$

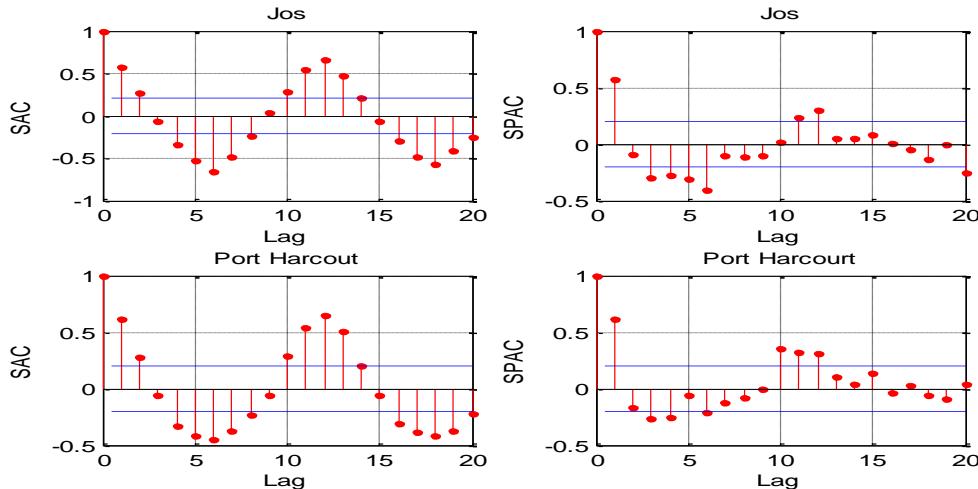
where,  $\hat{\emptyset}_k(\hat{e})$  are the autocorrelations of the residual  $\hat{e}$  from lag 1 to  $k$ . Since the alternative hypothesis is a two-tail test, we adopt the following decision rule:

$$\left. \begin{array}{l} \text{If } Q \geq \chi_{k-p-q, \alpha/2}^2 \text{ Reject } H_0 \\ \text{If } Q < \chi_{k-p-q, \alpha/2}^2 \text{ Not Reject } H_0 \end{array} \right\} \quad (18)$$

where,  $\chi_{k-p-q, \alpha/2}^2$  is the value from the  $\chi^2$  distribution table at half the preset significance level  $\alpha$  and  $k - p - q$  degrees of freedom. Consequently, a failure to reject the null hypothesis  $H_0$  suggests that the residual series is random and the model may be judged to be adequate. A complementary approach is by examining the autocorrelation plot of the residual series with a superimposed confidence bounds. The residual series is judged to be random if the autocorrelations at all lags fall within the confidence bounds.

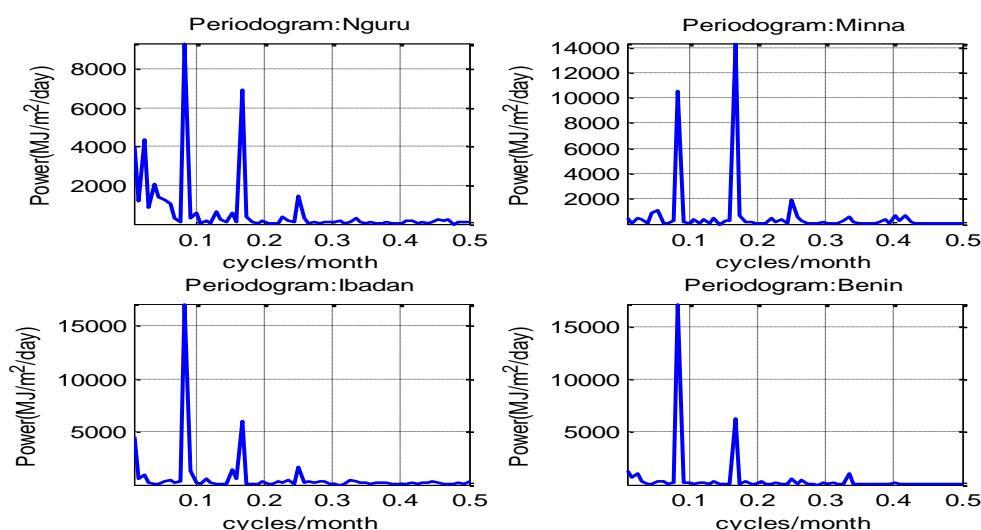
### 3.0. Results and Discussion

The correlograms for two stations (Jos and Port Harcourt) are as shown in Figure 1. It is evident that the correlograms show alternating spikes with repeated patterns and constant period indicating the presence of seasonal fluctuation. Harmonic analysis of the series showed that the first-two harmonics were significant for all the stations.



**Figure 1:** sample autocorrelation function (SACF) and sample partial autocorrelation function (SPACF) plots of monthly mean daily global solar radiation for Jos and Port Harcourt stations showing seasonal patterns

The spectral plots for Nguru, Minna, Ibadan and Benin are as shown in Figure 2. It can be seen from the spectral plots that there are two dominant frequencies at 0.083cpm and 0.17cpm accounting for most of the variability in the monthly mean daily global solar radiation at Sokoto and Ibadan. These frequencies were employed following the procedure outlined earlier in section 2.0 to determine the periodic (deterministic) component of the time series and from which the stochastic component was derived.

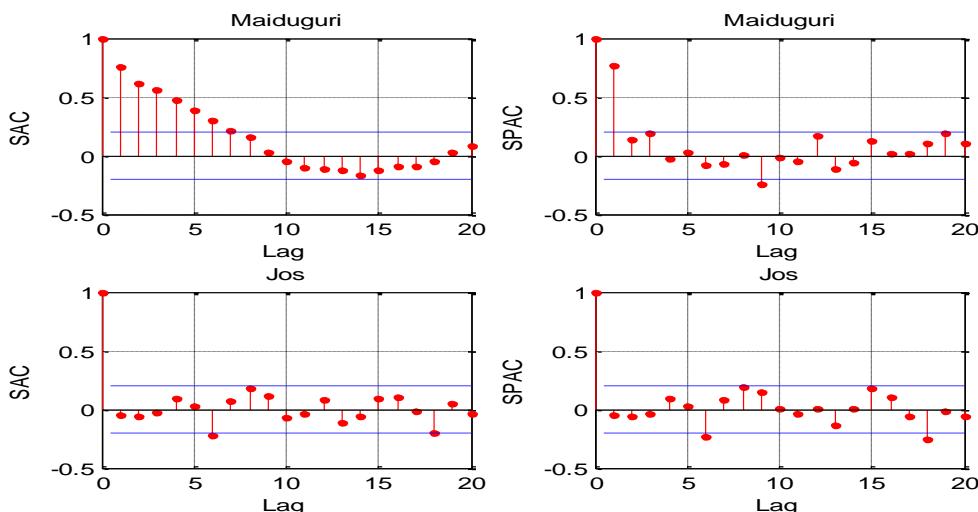


**Figure 2:** The spectral plots of the monthly mean daily global solar radiation for Nguru (a sub-sahelian station), Minna (a midland station), Ibadan (a savannah station) and Benin (a coastal station) stations with seasonal content showing two dominant frequencies at 0.083cpm and 0.17cpm; cpm = cycles per month

An inspection of the Phillips-Perron test statistic in Table 1 and the sample autocorrelation function plot of the modified series shown in Figure 3 indicate that the series was made stationary.

**Table 1:** Phillips-Perron univariate unit root test of the modified series

Station	pp value		C value $\alpha=0.05$
	Lag 0	Lag 1	
Sokoto	-4.08	-4.09	-3.85
Nguru	-4.34	-4.40	-3.85
Kano	-4.24	-4.19	-3.85
Maiduguri	-5.93	-4.95	-3.85
Bauchi	-4.26	-4.26	-3.85
Jos	-4.64	-4.64	-3.85
Minna	-4.55	-4.55	-3.85
Yola	-4.52	-4.65	-3.85
Bida	-4.28	-4.27	-3.85
Ibadan	-4.49	-4.49	-3.85
Ikeja	-4.83	-4.84	-3.85
Enugu	-4.47	-4.43	-3.85
Benin	-4.15	-4.14	-3.85
Port Harcourt	-5.25	-5.35	-3.85

**Figure 3:** sample autocorrelation function (SACF) and sample partial autocorrelation function (SPACF) plots of radiation residual for Maiduguri and Jos stations after removing seasonal component

However, resistance between monthly mean daily radiation anomalies (stochastic series) is seen to be significant at nine stations including Sokoto, Nguru, Kano, Maiduguri, Bauchi, Yola, Minna, Ibadan, and Benin. The other stations including Jos, Bida, Ikeja, Enugu and Port Harcourt showed insignificant autocorrelations at all lags. Figure 3 shows the correlogram for Maiduguri, a station with significant autocorrelation and the correlogram for Jos, a station with insignificant autocorrelation for illustration. For Maiduguri, the autocorrelations for the first few lags fell outside the 95% confidence bounds showing significant persistence whereas for Jos, the autocorrelation values at all lags fell within the 95% confidence bounds showing insignificant persistence in radiation anomalies at this station. This observation suggests that the conventional autoregressive model may be employed to forecast monthly mean daily global radiation anomalies at the nine stations where autocorrelation was found to be significant but not so for the other five stations with insignificant autocorrelation.

Following the method outlined in section 4.0 for identifying the structure and order of the model (e.g. see Figure 3 for Maiduguri); we tentatively identify the non-seasonal autoregressive model of order 1 given by:

$$r_t = \emptyset_1 r_{t-1} + e_t \quad (19)$$

The point estimate of the parameter  $\hat{\phi}_1$  in model (16) for the stations is as presented in Table 2. The statistical significance of model parameters was tested using t-test for the study stations and the result obtained is equally presented in Table 2. The fact that the t value absolute at Sokoto, Nguru, Kano, Maiduguri, Bauchi, Yola, Minna, Ibadan, and Benin are higher than  $t_{0.975}$  from the student's  $t$  distribution table causes us to reject the null hypothesis of zero autocorrelation at the preset significance level  $\alpha = 0.05$ .

**Table 2:** Study sites with their corresponding estimated model parameters and their t-statistic

Station	$\hat{\phi}_1$	Standard Error	T value absolute	$t_{0.975}$
Sokoto	0.4207	0.09533	4.413	1.98
Nguru	0.5712	0.08492	6.726	1.98
Kano	0.3218	0.09842	3.2697	1.98
Maiduguri	0.7615	0.06704	11.3589	1.98
Bauchi	0.2494	0.0999	2.4965	1.98
Jos	-0.0446	-0.1028	-0.4342	1.98
Minna	0.5492	0.08677	6.333	1.98
Yola	0.2740	0.09921	2.7618	1.98
Bida	-0.0415	-0.1016	-0.4087	1.98
Ibadan	0.5532	0.08537	6.48	1.98
Ikeja	0.1263	0.1026	1.231	1.98
Enugu	-0.0507	-0.1026	-0.4946	1.98
Benin	0.2496	0.09773	2.554	1.98
Port Harcourt	0.1289	0.0999	1.2903	1.98

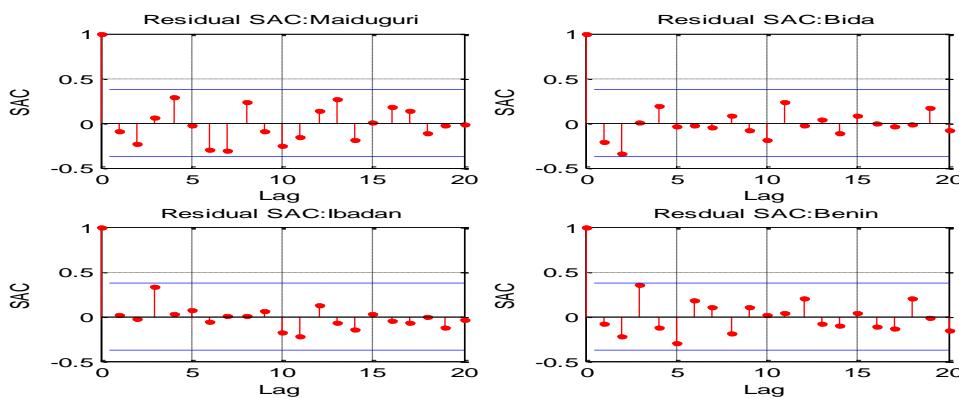
Thus we can conclude that the estimated parameters at those stations are statistically significant. However, a comparison of the t value absolute of parameters at Jos, Bida, Ikeja, Enugu and Port Harcourt and  $t_{0.975}$  causes us not to reject the null hypothesis of zero autocorrelation since the computed  $t$  statistic at those stations are lower than  $t_{0.975}$ . Thus, we would conclude that the parameter estimates at those stations are not statistically significant.

The Ljung-Box statistic of the error series was computed for all the study stations and the result is presented in Table 3. The computed Q statistic was found to be lower than the critical value  $\chi^2_{0.975}$  from the  $\chi^2$  distribution table at the preset significance level of 0.05 and 19 degrees of freedom. This causes us not to reject the null hypothesis of zero autocorrelation or randomness in the error series. In addition, the autocorrelation function (ACF) of the model residuals plotted for all the station revealed that the residuals are random with the autocorrelations falling within the 95% confidence interval around zero.

**Table 3:** Ljung-Box statistic of model residual series for all study stations

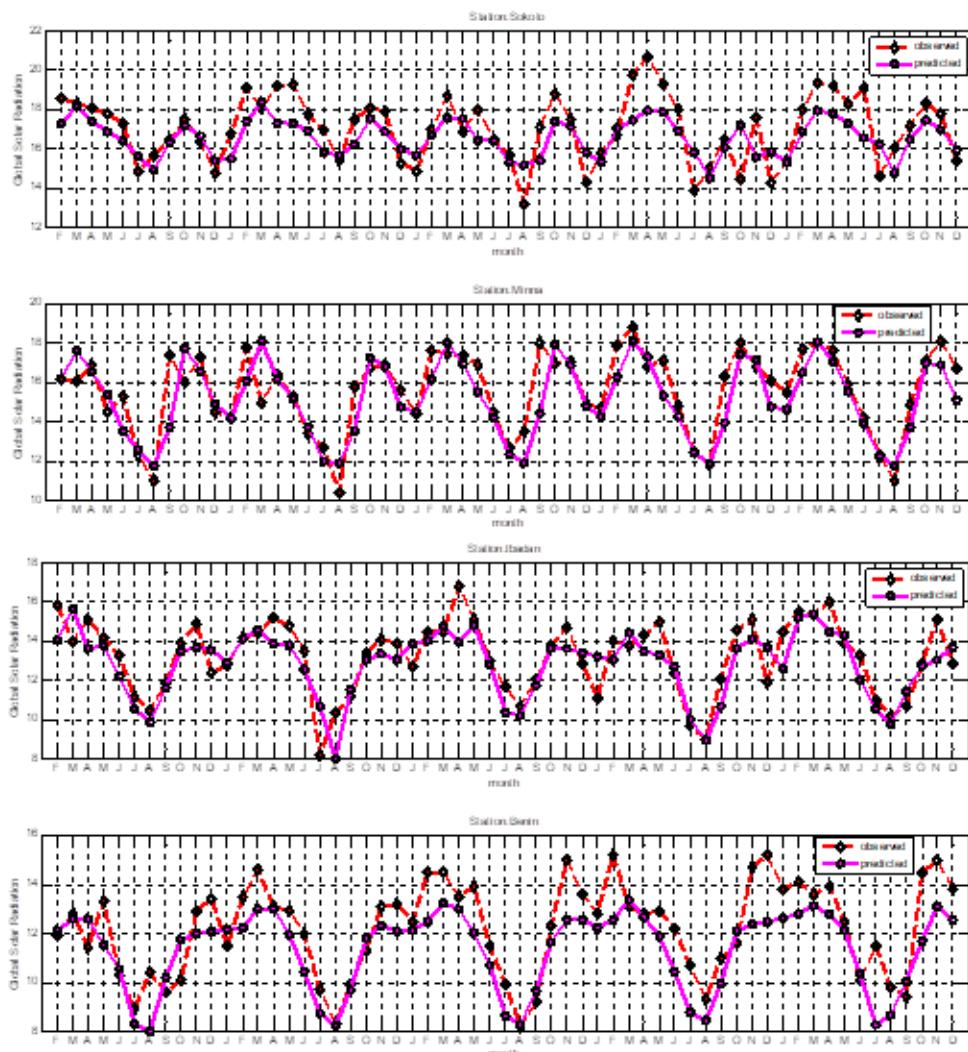
Station	$Q$ value	$\chi^2_{0.975}$	$df$	upto lag	$\alpha$
Sokoto	21.34397	32.85233	19	20	0.05
Nguru	26.22795	32.85233	19	20	0.05
Kano	19.6778	32.85233	19	20	0.05
Maiduguri	22.96371	32.85233	19	20	0.05
Bauchi	32.22676	32.85233	19	20	0.05
Jos	14.14422	32.85233	19	20	0.05
Minna	32.22676	32.85233	19	20	0.05
Yola	32.22676	32.85233	19	20	0.05
Bida	15.631	32.85233	19	20	0.05
Ibadan	20.24178	32.85233	19	20	0.05
Ikeja	17.03581	32.85233	19	20	0.05
Enugu	28.88052	32.85233	19	20	0.05
Benin	19.92759	32.85233	19	20	0.05
Port Harcourt	26.96986	32.85233	19	20	0.05

Figure 4 illustrates the residual autocorrelation plots for Maiduguri, Bida, Ibadan and Benin. This is the desired result as effective autoregressive model should explain persistence and yield random residuals (Chatfield, 2004). Thus, we can conclude that at the preset significance level of 0.05, we cannot reject the adequacy of the autoregressive models for forecasting purposes.



**Figure 4:** Model residual autocorrelation function (RSACF) plots for Maiduguri (a sub-sahelian station), Bida (a midland station), Ibadan (a Guinea Savaanah station) and Benin (a coastal station) stations showing zero autocorrelation of model errors.

The plots of Figure 5 illustrates a great deal of agreement between predicted and measured values of the monthly mean global solar radiation at Sokoto (a Sub-Saharan station), Minna (a Midland station), Ibadan (a Guinea savannah station) and Benin (a Coastal station) indicating that the observed data is well monitored by the model for those stations. However, errors of intensities are common especially around turning points.



**Figure 5:** Simulation of monthly mean daily Global solar radiation for Sokoto (a sub-sahelian station), Minna (a midland station), Ibadan (a Guinea savannah station) and Benin (a coastal station) stations.

This observation could be due to the fact that the model did not take into account the effect of several weather phenomena such as the cloud cover and varying atmospheric constituents in estimating the global solar radiation. The performance of the model was evaluated by computing the coefficient of determination ( $R^2$ ), the mean bias error (MBE), and the root mean square error (RMSE) of prediction and the result is presented in Table 4.

**Table 4:** Result of prediction statistics showing the coefficient of determination ( $R^2$ ), the mean bias error (MBE) and the root mean square error (RMSE) of predictions

Station	$R^2$	MBE	RMSE
Sokoto	0.4715	0.347	0.775
Nguru	0.6847	0.126	0.564
Kano	0.5636	-0.124	0.675
Maiduguri	0.7968	0.248	0.846
Bauchi	0.3324	0.000	0.954
Jos	0.3339	0.844	1.385
Minna	0.5522	-0.299	0.847
Yola	0.8692	0.138	1.092
Bida	0.2918	0.344	1.023
Ibadan	0.4346	0.346	0.899
Ikeja	0.1661	-1.418	1.859
Enugu	0.3209	-0.058	0.843
Benin	0.2475	0.593	1.145
Port Harcourt	0.1686	0.642	1.284

Generally, the model gave a relatively high  $R^2$  values and low RMSE especially for stations within the sub-sahelian region. However, similar performance is seen at Minna (55.22%) and Yola(86.92%) in the midland region as well as Ibadan(43.46%); a Guinea savannah region. The situation in the coastal region is relatively different with poorer performance as indicated with lower  $R^2$  and higher RMSE values. This tends to suggest that the model is region selective, performing best in the sub-sahelian region but very poor in the coastal region. The MBE values showed that the model estimates were slightly underestimated at Kano, Minna, Ikeja and Enugu but also slightly overestimated for most of the other stations studied.

#### 4.0. Conclusions

The solar radiation data taken from 14 meteorological stations in Nigeria has been utilized to develop self-predicting models for monthly mean daily global solar radiation forecast over Nigeria. After removing the deterministic component through harmonic analysis, it was found that the stochastic component of the monthly mean daily global solar radiation over most of the study locations may well be monitored by the first-order autoregressive model, AR(1). The situation at Jos, Bida, Ikeja, Enugu and Port Harcourt however suggests that the underlying mechanism generating the stochastic series at those stations is not autoregressive. In addition, a simulation of the measured global solar radiation revealed that the model did well for the stations within the sub-sahelian region but not so well for the stations in the coastal region. Hence, the model is recommended for use to forecast solar radiation over all the stations within the sub-sahelian region but not appropriate for the coastal stations. The situation in the midland region showed that the model is station specific in terms of performance whereas it is difficult to conclude on the Guinea savannah region with result from only one station.

#### References

Adamowski, K. and Smith, A. F. (1972). Stochastic generation of rainfall. *Journal of Hydraulics Division, American Society of Civil Engineers*, 98(11), pp. 1935-1945.

Adeyemi, B. and Fasonranbaku, A. O. (2008). Forecasting of daily mean air temperature over Nigeria using stochastic processes. *Science Research Annals*, 4(2), pp. 6-14.

Akpabio, L. E., Udo, S. O. and Etuk, S. E. (2004). Empirical correlation of global solar radiation with meteorological data from Onne, Nigeria. *Turkish Journal of Physics*, 28(3), pp. 301-307.

Barnett, T. P., Heinz, H. D. and Hasselmann, K. (1984). Statistical prediction of seasonal air temperature over Eurasia. *Tellus A: Dynamic Meteorology and Oceanography*, 36(2), pp. 132-146.

Bowerman, B. L., O'Connell, R. T. and Koehler, A. B. (2005). Forecasting, time series, and regression: An applied approach. 4<sup>th</sup> Ed., Belmont, CA: Thomson Learning.

Box, G. E. P. and Jenkins, G. M. (1976). Time series analysis: Forecasting and control. Holden-Day, San Francisco

Brooks, D. R. (2006). Monitoring solar radiation and its transmission through the atmosphere. The GLOBE program's Aerosols, water vapor, and UV-A Monitoring Projects, V 2.0

Chatfield, C. (2004). The analysis of time series: An introduction. Chapman and Hall, CRC Press, Boca Raton.

Durbin, J. (1960). The fitting of time series model. *Review of the International Statistical Institute*, 28(3), pp. 233-244.

Falayi, E. O. and Rabiu, A. B. (2005). Modeling global solar radiation using sunshine duration data. *Nigerian Journal of Physics*, 178, pp. 181-186.

Fuller, W. A. (1976). Introduction to statistical time series. John Wiley and Sons, New York. P373. ISBN 0-471-28715-6.

Jones, R. N. (1964). Spectral analysis and linear prediction of meteorological time series. *Journal of Applied Meteorology and Climatology*, 3(1), pp. 45-52.

Lindsey, R. (2009). Climate and Earth's energy budget: Feature articles, Earth observatory, part of the EOS project science office, NASA Goddard Space Flight Center. p. 4

Ljung, G. M. and Box, G. E. P. (1978). On a measure of a lack of fit in time series models. *Biometrika*, 65(2), pp. 297-303.

Okundamiya, M. S. and Nzeako, A. N. (2011). Empirical model for estimating global solar radiation on horizontal surfaces for selected cities in the six geopolitical zones in Nigeria. *Journal of Control Sciences and Engineering*, <http://dx.doi.org/10.1155/2011/356405>.

Page, J. K. (1964). The estimation of monthly mean values of daily total shortwave solar radiation on vertical and inclined surfaces from sunshine records for latitude 40°N - 40°S. *Proceedings of the UN Conference on New Sources of Energy*, 4, pp. 378.

Raji, T. A., Boyo, A. O. and Mehdi, H. (2012). Analysis of global solar radiation data as time series data for some selected cities of western part of Nigeria. *International Journal of Advanced Renewable Energy Researches*, 1(1), pp. 14-19.

Safari, B. and Gasore, J., (2009). Estimation of global solar radiation on horizontal surface in Rwanda using empirical model. *Asian Journal of Scientific Research*, 2(2), pp. 68-75.

Sharma, T. C. (1989). Stochastic modeling of rainfall processes in the Central African tropics. *Atmospheric Deposition (Proceedings of the Baltimore Symposium)* IAHS Publication No. 179.

Smadi, A. A. (2009). Periodic autoregression modeling of the temperature data of Jordan. *Pakistan Journal of Statistics*, 25(3), pp. 323-332.

Sulaiman, M. Y., Hlaing Oo, W. M., Abd. Wahab, M. and Sulaiman, Z. A. (1997). Analysis of residuals in daily solar radiation time series. *Renewable Energy*, 11, pp. 97- 105.

Tesfaye, Y. G., Meerschaert, M. M. and Anderson, P. L. (2006). Identification of periodic autoregressive moving average models and their application to the modeling of river flows. *Water Resources Research*, 42(1), W01419, doi: 10.1029/2004WR003772.

Trabea, A. A. and Shaltout, M. M. (2000). Correlation of global solar radiation with meteorological parameters over Egypt. *Renewable Energy*, 21(2), pp. 297- 308.

Zaharim, A., Razali, M. A., Gim, T. P. and Sopian, K. (2009). Time series analysis of solar radiation data in the tropics. *European Journal of Scientific Research*, 25(4): 672- 678.

Zakaria, A. (2011). A study of periodic and stochastic modeling of monthly rainfall from Purajaya station. *Asian Transaction on Engineering* , 01(03), pp. 1-7.

---

**Cite this article as:**

Ocholi M., Adeyemi, B., Omojola O. O. and Samuel C. S. 2021. Evaluating the Performance of Autoregressive Model for Solar Radiation Forecasting. *Nigerian Journal of Environmental Sciences and Technology*, 5(2), pp. 444-455. <https://doi.org/10.36263/njest.2021.01.0289>